

## IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Liam D. Comerford  
Serial No.: 10/674,573  
Filed: September 30, 2003  
For: DISSECTION OF UTTERANCES INTO  
COMMANDS AND VOICE DATA

Examiner: Lennox, N.  
Group Art Unit: 2626

**Pre-Appeal Brief Request for Review**

This is a Pre-Appeal Brief Request for Review, in response to the Final Office Action mailed December 19, 2007.

Applicants respectfully contend that the claim rejections set forth in the Final Office Action are clearly erroneous for at least those reasons set forth in the Amendment filed on October 9, 2007, which is incorporated herein by reference. Rather than reiterate those reasons in full, the following will focus on the errors in the Examiner rejections of independent Claims 1, 16, and 32.

**Rejections under 35 U.S.C. 102(e)**

Independent Claims 1 and 16 stand rejected under 35 U.S.C. §102(e) as being anticipated by U.S. Patent Publication No. 2005/0071169 to Steinbiss (hereinafter "Steinbiss"). Steinbiss does not disclose the following limitations of Claim 1:

decoding at least one word in acoustic data representing an acoustic signal that comprises a human utterance and determining acoustic word boundaries within the acoustic data; ... and

identifying acoustic data segments in the utterance based on the acoustic word boundaries.

Moreover, Steinbiss does not disclose the following limitations of Claim 16:

decoding at least one word in voice data representing the acoustic signal that comprises a human utterance and determining the acoustic word boundaries within the voice data; ... and

associating segments in the voice data based on the acoustic word

boundaries with labels.

With respect to, *inter alia*, the above-recited limitations of Claims 1 and 16 relating to determining/recognizing acoustic word boundaries, the Examiner has cited Steinbiss and stated the following “Fig. 1 illustrates voice command S with word sequence ‘TV on’, where signal section tl represents the word ‘tv’ and signal section tr represents the word ‘on’” (Office Action, pp. 4-5).

However, Steinbiss does not show “determining/recognizing acoustic word boundaries” as essentially recited in Claims 1 and 16. This is evident from each of the figures in Steinbiss, namely Figures 1, 2, and 3, which each simply show the voice signal S, but not any of the word boundaries for the words (i.e., “TV” and “on”) comprised in the voice signal S.

What Steinbiss does disclose, with respect to the voice signal S, is a start time instant t1 for the entire voice signal S and an end time instant t2 for the entire voice signal S. Moreover, Steinbiss discloses a reference time instant tr. However, as explicitly disclosed in Steinbiss, “the end of the vowel ‘o’ in the word ‘on’ is the reference point in this connection” (Steinbiss, paragraph [0038]). Moreover, as explicitly disclosed in Steinbiss, in a variant, “the end of the voice signal S is simply chosen as reference time instant tr” (Steinbiss, paragraph [0040]). Thus, with respect to the preceding disclosed variant of Steinbiss, Steinbiss explicitly defines tr NOT as the end of the last word in the word signal, but rather as the end of the voice signal. This explicit definition by Steinbiss is clearly a signal boundary.

However, neither the end of a vowel in the middle of a word, nor the end of the voice signal S (i.e., a signal boundary), is equivalent to acoustic word boundaries as recited in each of Claims 1 and 16. For example, as disclosed at paragraph [0063] of the Applicants’ specification, the beginning and ending of a voice signal S may actually correspond to the closing and opening of a microphone button and not to any words. In that example provided in the Applicants’ specification with respect to Figure 4, “elements 401 and 410 represent the beginning and ending of the buffered acoustic data, ... and correspond to closing 401 and opening 410 a microphone button (e.g., a physical switch)”.

Steinbiss does not separate commands from acoustic data as contemplated by the present claims. In fact, Steinbiss provides a system that is limited to only recognizing commands.

In contrast, we process a same utterance for both “command” words (known to be directly acted upon) and “non-command” words (i.e., “acoustic data”, as explicitly and differentially recited (with respect to “command”) in Claims 1, 16, and 32). Hence, the separate recitations in these claims of “commands” and “acoustic segments”, the latter

corresponding to the “acoustic data”. In addition to identifying a command from a particular utterance, we further process the same utterance to obtain acoustic segments using the acoustic word boundaries. This is different from the prior art, which does each feature separately, i.e., command (including command sequence) processing and non-command (acoustic data) processing. Steinbiss explicitly DEFINES the ENTIRE PHRASE “TV on” as simply a “COMMAND SEQUENCE” (see, e.g., Steinbiss, paragraph [0039]). That is, the sequence of words makes up a single command. However, simply determining commands, even when such commands include more than one word, is not the dual functionality provided by Claims 1 and 16 regarding extracting commands and acoustic data in a same utterance, with all the attendant advantages provided thereby.

It is not surprising that Steinbiss does not disclose “determining/recognizing acoustic word boundaries” as essentially recited in Claims 1 and 16, since the reference time instant  $t_r$  is not intended to have any relation to word boundaries, but rather depends “on the occurrence and/or time variation of the voice signal” irrespective of acoustic word boundaries per se (see, e.g., Steinbiss, Abstract). This is because the “switch-on action A does not, however, take place immediately after the recognition of the command sequence by the voice recognition device, but only at a defined action time instant  $t_a$  that is at a fixed time interval  $\Delta a$  with respect to the reference time instant” (Steinbiss, paragraph [0039]).

Hence, it is clear that Steinbiss does not disclose “determining/recognizing word boundaries” as essentially recited in Claims 1 and 16.

With respect to identifying acoustic data segments based on the acoustic word boundaries as recited in Claim 1, and associating segments in the voice data based on the acoustic word boundaries with labels as recited in Claim 16, clearly Steinbiss cannot disclose the same when Steinbiss does not even disclose “determining/recognizing word boundaries” in the first place.

It is to be appreciated that the Examiner has stated the following on page 5 of the Office Action with respect to the preceding limitations of Claims 1 and 16 relating to identifying/associating segments: “Fig. 1, acoustic data segments  $t_l$  and  $t_r$ ”.

However, as argued above,  $t_l$  is simply the start time instant for the entire voice signal S (i.e., a signal boundary) and  $t_r$  is, for example, either “the end of the vowel ‘o’ in the word ‘on’” (Steinbiss, paragraph [0038]) or “the end of the voice signal S” (Steinbiss, paragraph [0040]) (i.e., a signal boundary).

Hence, no word boundaries are determined/recognized and then used in Steinbiss, as recited in Claims 1 and 16.

“A claim is anticipated only if each and every element as set forth in the claim is found, either expressly or inherently described, in a single prior art reference.” MPEP §2131, citing *Verdegaal Bros. v. Union Oil Co. of California*, 814 F.2d 628, 631, 2 USPQ2d 1051, 1053 (Fed. Cir. 1987).

Accordingly, Claims 1 and 16 are patentably distinct and non-obvious over Steinbiss for at least the reasons set forth above.

**Rejections under 35 U.S.C. 103(a)**

Independent Claim 32 stands rejected under 35 U.S.C. §103(a) as being unpatentable over Steinbiss in view of U.S. Patent Publication No. 2002/0111803 to Romero (hereinafter “Romero”).

It is respectfully asserted that none of the cited references disclose the following limitations of Claim 32 (emphasis added):

a system for recognizing commands and voice data in a same utterance including ... at least one program that executes label-identified commands and processes remaining portions of the utterance including processing audio data parts separately from the commands using a different vocabulary, the vocabulary being selected in accordance with at least one command in the utterance.

While the Examiner has cited Walker as disclosing the same, the cited portions of Walker (e.g.: col. 3, lines 12-25; col. 4, lines 34-37; col. 4, lines 43-49) relate to label-identified commands, and not remaining portions of the utterance and, hence, do not disclose all of the limitations recited in Claim 32. That is, the feature/value table shown at column 3, lines 3-7 of Walker is further described at column 3, lines 12-25 of Walker (which was cited by the Examiner). However, this feature/value table is formed from the results of the TAG parser program that parses (command) tags (see, e.g., Walker, col. 2, lines 53-67). The same is true for the other sections of Walker cited by the Examiner. For example, the Examiner has cited column 4, lines 34-37 of Walker, which simply relates to an application program being directly referenced from (scripting language within) the tags defined by the rule grammar (see, e.g., Walker, col. 4, lines 29-31), where the cited column 4, lines 34-37 specifically and simply show a portion of the rule grammar for such an embodiment of Walker. Again, with respect to the column 4, lines 43-49, the “tags parser program is invoked”. The tags are used for commands (see, e.g., Walker, col. 1, line 67 to col. 2, line 9). Thus, in the cited sections of Walker, Walker is further processing commands, but not the remaining portions of the utterance including processing audio data parts separately from the commands, let alone using a different

vocabulary, let alone with the vocabulary being selected in accordance with at least one command in the utterance, all as recited in Claim 32.

Hence, the cited combination of Walker and Romero do not disclose all the limitations recited in Claim 32.

"To establish prima facie obviousness of a claimed invention, all the claim limitations must be taught or suggested by the prior art" (MPEP §2143.03, citing *In re Royka*, 490 F.2d 981, 180 USPQ 580 (CCPA 1974)).

Accordingly, Claim 32 is patentably distinct and non-obvious over Steinbiss for at least the reasons set forth above.

Regarding the rejection of the remaining dependent claims under both 35 U.S.C. 102(e) and 35 U.S.C. 103(a), dependent Claims 2-15, 17-31, and 33-36 directly or directly depend from Claims 1, 16, and 32, respectively, and, thus, include all the limitations of Claims 1, 16, and 32, respectively. Thus, Claims 2-15, 17-31, and 33-36 are patentably distinct and non-obvious over the cited references for at least the reasons set forth above with respect to Claims 1, 16, and 32, respectively.

Reconsideration of the rejections and allowance of the pending claims is earnestly solicited.

Respectfully submitted,

Dated: March 5, 2008

By: Gaspard J. Randazzo  
Gaspard J. Randazzo  
Registration No. 41,528

**Mailing Address:**

**KEUSEY, TUTUNJIAN & BITETTO, P.C.**

**20 Crossways Park North, Suite 210**

**Woodbury, NY 11797**

**Tel: (516) 496-3868**

**Fax: (516) 496-3869**